

SYLLABUS

Subject: Data Mining Methodology II (STA6704)
Lecture: M & W 6:00-7:15 pm, Room 220, CL1
Instruction Period 01/07/2008 – 04/21/2008
Holidays 1/21/2008, Monday, Martin Luther King Jr.’s Day
 3/10/2008 - 3/15/2008, Spring Break
Instructor: Dr. Xiaogang Su
 Room 102, CC II
 (407) 823-2940 [O]
 xsu@pegasus.cc.ucf.edu
Office Hours: W 3:00- 4:00pm and R 2:00-3:00pm
Text Book: T. Hastie, R. Tibshirani, and J. Friedman (2001)
The Elements of Statistical Learning – Data Mining, Inference, and Prediction, Springer Series in Statistics, NY
Course Web Page: <http://pegasus.cc.ucf.edu/~xsu/CLASS/STA6704>
Prerequisite: Linear Models, Regression Analysis, and Knowledge of Basic Data Mining Techniques.

- **Course Objective:** As the second course in the data mining sequence, this course focuses mainly on neural network modelling. SAS will be used intensively for this course. In this class, you will learn to use a set of data mining tools by applying them to real problems from business world. At the end of the course, you are expected to have solid knowledge about classic and modern data mining techniques and hands-on experience in using SAS to perform data mining tasks. Some SAS macro programs will be discussed. I will also introduce the free software R and discuss the implementation of various data mining techniques in R.

Both Topics in unsupervised and supervised prediction are to be covered. A more specific list includes neural network architecture, foundation of linear and (parametric/nonparametric) nonlinear regression modelling, generalized additive models (GAM), Support vector machine (SVM), multivariate adaptive regression splines (MARS), recursive partitioning and its extensions (e.g., bagging, boosting, and random forests).

- **Grading:** There will be a number of homework assignments, in-class quizzes, and a final project during the semester.

For the final project, students are given the freedom to select data from whatever field they are interested in (and at the same time, a topic that is adequate for the course). Students should make their own plans to collect data, raise some interesting questions to study, and consult me for the adequacy of the project. Also, each student will have the opportunity to present their work in class.

The grade of this course is based on the performance of projects: The homework assignments are worth 75 points in total, the final project is 20 points, and in-class quizzes make up 15 more points. The grade is given according to the following scale. Note that the total is 110.

Range	94+	93-90	89-87	86-83	82-80	79-77	76-73	72-70	69-67	66-63	62-60	59-0
Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F